Intelligente Persönliche Assistenten im häuslichen Umfeld

Erkenntnisse aus einer linguistischen Pilotstudie zur Erhebung audiovisueller Interaktionsdaten

Tim Moritz Hector & Christine Hrncal *Universität Siegen*

WORKING PAPER SERIES | NO. 14 | MÄRZ 2020

Collaborative Research Center 1187 Media of Cooperation
Sonderforschungsbereich 1187 Medien der Kooperation

Working Paper Series
Collaborative Research Center 1187 Media of Cooperation

Print-ISSN 2567–2509
Online-ISSN 2567–2517
DOI https://doi.org/10.25819/ubsi/1013
Handle https://dspace.ub.uni-siegen.de/handle/ubsi/1573
URN urn:nbn:de:hbz:467-15731

Publication of the series is funded by the German Research Foundation (DFG).

This Working Paper Series is edited by the Collaborative Research Center Media of Cooperation and serves as a platform to circulate work in progress or preprints in order to encourage the exchange of ideas. Please contact the authors if you have any questions or comments. Copyright remains with the authors.

The Working Papers are accessible via the website http://wp-series.mediacoop.uni-siegen.de or can be ordered in print by sending an email to workingpaperseries@sfb1187.uni-siegen.de

Universität Siegen
SFB 1187 Medien der Kooperation
Herrengarten 3
57072 Siegen, Germany
www.sfb1187.uni-siegen.de
workingpaperseries@sfb1187.uni-siegen.de

Intelligente Persönliche Assistenten im häuslichen Umfeld

Erkenntnisse aus einer linguistischen Pilotstudie zur Erhebung audiovisueller Interaktionsdaten

Tim Moritz Hector & Christine Hrncal *Universität Siegen* — tim.hector@uni-siegen.de

Abstract Sprachassistenten werden in einer steigenden Zahl von Haushalten in den Alltag eingebunden. Es zeigen sich dabei sprachliche und kulturelle Praktiken, die durch die Integration artifizieller Mündlichkeit in die Interaktion entstehen, wie sie bisher noch nicht beschrieben werden konnten. Diese untersucht der gesprächslinguistisch ausgerichtete Teilbereich des Projekts B06 „Un/erbetene Beobachtung in Interaktion: ‚Intelligente Persönliche Assistenten' (IPA)" im Sonderforschungsbereich „Medien der Kooperation" an der Universität Siegen. Sprachassistenzsysteme sind außerdem für ihre Funktionalität auf die dauerhafte Beobachtung des häuslichen Umfelds angewiesen. Die Reflexion der NutzerInnen über dieses „Mithören", das im öffentlichen Diskurs teilweise sehr kritisch betrachtet wird, steht ebenfalls im Fokus der im Projekt durchgeführten Untersuchungen. Im Rahmen der hier vorgestellten Pilotstudie werden methodische Prämissen im Hinblick auf das Vorgehen bei der Datenerhebung reflektiert und aus den gewonnenen Daten erste Anhaltspunkte für die sprachwissenschaftlichen Analysekategorien herauskristallisiert. Der Schwerpunkt liegt dabei auf der Identifikation von sprachlich-interaktionalen Praktiken und deren Einbettung in soziokulturelle Praktiken, die in der Hauptstudie ebenfalls näher beleuchtet werden sollen. Unsere Daten zeigen, dass Interagierende ein Sprachassistenzsystem nicht wie einen zusätzlichen Gesprächsteilnehmer in die Interaktion einbeziehen, sondern es durchaus wie ein technisches Gerät behandeln. Gleichzeitig scheint die parallele Nutzung des medial mündlichen Kanals zur Bedienung eines Geräts auf der einen und zum Führen einer Konversation auf der anderen Seite Auswirkungen auf das Repertoire sprachlich-interaktionaler sowie kultureller Praktiken zu haben.

Keywords Intelligente Persönliche Assistenten, Voice User Interfaces, Gesprächsanalyse, Beobachtung, Mensch-Maschine-Interaktion

Die im Folgenden beschriebene Untersuchung versteht sich als Pilotstudie für das Projekt „Un/erbetene Beobachtung in Interaktion: ‚Intelligente Persönliche Assistenten' (IPA)". Das Projekt ist empirisch angelegt und basiert auf im häuslichen Umfeld erhobenen Audio- und Videoaufzeichnungen. Dadurch entsteht eine „doppelte" Beobachtung auf der Meta-Ebene: Die Beobachtung durch die Intelligenten Persönlichen Assistenten wird durch die zusätzliche Aufzeichnung des Geschehens durch ForscherInnen ebenfalls dokumentiert – und möglicherweise beeinflusst. Um u.a. mögliche Verfahren für den Umgang mit diesem „Beobachterparadoxon" (vgl. Labov 1972) zu erproben, entsprechende Datenquellen zu identifizieren sowie mögliche Fehlerquellen bei der Datenerhebung zu entdecken und ferner erste Befunde zu näher zu betrachtenden sprachlichen Praktiken gewinnen zu können, wurde eine Pilotstudie durchgeführt. In dieser wurden in Vorbereitung auf die Hauptstudie drei Fragestellungen

näher verfolgt: (1) Welche Quellen zur Gewinnung von Daten erweisen sich als ergiebig? Wie hoch ist die Bereitschaft von angesprochenen potenziellen ProbandInnen, als StudienteilnehmerInnen zur Verfügung zu stehen und wie kann diese ggf. erhöht werden? (2) Mit welchen Herausforderungen ist bei der Datenerhebung zu rechnen, welche Fehlerquellen können identifiziert werden? Dabei wird insbesondere betrachtet, wie mit der ‚doppelten Beobachtung' umgegangen werden kann. Zudem wurde geprüft, ob die entwickelten Instruktionen für die aufgenommenen Haushalte sich eignen, um diese hinreichend auf die Aufnahmesituation vorzubereiten. (3) Welche sprachlich-interaktionalen Praktiken zeigen sich im Zusammenhang mit der Nutzung von häuslichen Sprachassistenzsystemen? Im Hinblick auf diese Frage rücken sowohl Mensch-Maschine-Interaktionen wie auch Mensch-Mensch-Interaktionen in den Fokus. Die Voruntersuchung zielt auf die Identifikation von Praktiken, die sich für eine genauere Betrachtung in der Hauptstudie anbieten. Dabei wird der Blick hier zunächst auf sprachlich-interaktionale Praktiken im konversationsanalytischen bzw. interaktional-linguistischen Sinne gerichtet (vgl. Schegloff 1997; Selting 2016). Daran anschließend wird in der Hauptstudie auch die Beziehung dieser interaktionalen „Infrastruktur" (Schegloff 2012) zu „kulturellen Verstehenshintergründen", d.h. soziokulturellen Praktiken im Sinne Schatzkis (2002) in den Blick genommen, in die das individuelle Handeln eingebettet ist (vgl. Habscheid 2016, S. 133).

Das Vorgehen im Projekt ist im Sinne der ethnomethodologischen Konversationsanalyse datengeleitet und die Daten werden ohne spezifische Vorannahmen über mögliche sprachliche Formen und Strukturen untersucht: „Jedes Textelement wird zunächst einmal – auch wenn dies ganz unwahrscheinlich erscheinen mag – als Bestandteil einer sich reproduzierenden Ordnung betrachtet und in den Kreis möglicher und relevanter Untersuchungsphänomene einbezogen" (Bergmann 2001, S. 923). Um den Kreis der möglicherweise relevanten Untersuchungsphänomene enger fassen zu können und um in der späteren Hauptstudie Anhaltspunkte zu gewinnen, auf welche möglicherweise musterhaft auftretenden Phänomene zu achten ist, soll das aus der Pilotstudie gewonnene Datenkorpus erste Eindrücke vom Umgang mit den Sprachassistenzsystemen vermitteln, zu denen bislang nur wenige empirische, linguistische Studien vorliegen. Auf diese Weise stehen die methodologischen Prinzipien, nach denen in der Untersuchung im Projekt vorgegangen werden soll, auch im Einklang mit der Interaktionalen Linguistik, die als eines ihrer Leitprinzipien die Entwicklung der Analysekategorien „aus den Daten heraus" vorsieht (Selting und Couper-Kuhlen 2001, S. 277).

Die Sprachdaten konnten in verschiedenen Kontexten erhoben werden, zu denen auf unterschiedliche Weise ein Zugang entwickelt wurde: Einerseits konnte ein Feldzugang über Aufrufe in thematisch passenden Seminaren und durch Datenspenden ermöglicht werden, andererseits wurden im privaten Umfeld der AutorInnen potenzielle SpenderInnen identifiziert. Im Fokus standen dabei die Situationstypen bzw. Erhebungszeitpunkte, die auch in der Hauptstudie von Interesse sein werden, d.h. die audiovisuell festzuhaltende Ersteinrichtung des IPA, die per Audioaufnahme aufzuzeichnenden „Besuchssituationen" und die damit einhergehende Vorführung des IPA. Darüber hinaus wurde die routinierte Nutzung als zusätzlicher Datentyp hinzugezogen, der Aufschluss über bereits verfestigte sprachliche Muster liefert.

Methodologische Einordnung

Im Rahmen des Projekts wird u.a. mit einer Form der ethnographischen Gesprächsanalyse gearbeitet, die den Grundzügen der ethnomethodologischen Konversationsanalyse (KA) ähnelt, wie sie von Harold Garfinkel und Harvey Sacks entwickelt wurde. Diese fragt danach, wie sich Interagierende im Gespräch Sinn und Ordnung ihrer (sprachlichen) Handlungen gegenseitig aufzeigen (vgl. Bergmann 2001; Sacks 1992). Die „analytische Mentalität" (Schenkein 1978) der Konversationsanalyse arbeitet u.a. mit einem streng naturalistischen Empiriebegriff: Die AnalytikerInnen leiten ihre Schlussfolgerungen ausschließlich aus den Audio-Aufnahmen natürlicher, gesprochen-sprachlicher Interaktionen ab und beziehen keine weiteren Informationen über die SprecherInnen ein. Ausschließlich die sprachlich ausgedrückten Dokumentationen von Verständnis bzw. „Aufzeigeleistungen" (Deppermann 2008, S. 86) der Interagierenden untereinander, d.h. wie die InteraktantInnen selbst das Handeln des Gegenübers ‚verstehen', ist Gegenstand der Analyse, da diese Dokumentation für das Gegenüber gleichsam für die ForscherInnen das Verständnis der Handlungen zugänglich macht. Eine strikte Orientierung an diesem naturalistischen Empiriebegriff verliert allerdings die kulturellen Wissenshintergründe, an denen sich die Beteiligten bei der wechselseitigen Verfertigung gemeinsamer Abläufe (vgl. Schüttpelz und Meyer 2017, S. 158) orientieren, aus dem Blick. Mit Deppermann (2000) gehen wir deshalb davon aus, dass das „Verstehen" in der Interaktion keine Frage des „bloßen Ablesens" ist, sondern auch Hintergrundinformationen etwa zur Gesprächshistorie, zur Beziehung der Interagierenden sowie zum Kontext der Interaktion eine gewichtige Rolle spielen. Ausgehend vom erhobenen Datenmaterial als wichtigster Bezugspunkt (vgl. Deppermann 2000) rückt bei der Rekonstruktion und Interpretation des Handelns der Beteiligten das von ihnen nicht explizierte Wissen, das sie

im Gespräch voraussetzen und relevant machen, als wichtige Deutungsressource in den Fokus.

Anspruch / Beobachterparadoxon

Mit den oben dargestellten Prämissen verbunden ist in der Sprachwissenschaft stets die Abgrenzung von strukturalistischen Ansätzen, die Sprache losgelöst von ihrem Kontext betrachten. Gleichzeitig wendet sich die Forderung nach natürlichen Daten gegen experimentelle Settings und gegen elizitierte Daten, die nicht der realen Lebenswelt der SprecherInnen entsprechen (vgl. Gerwinski und Linz 2018, S. 108). Daraus leiten sich Ansprüche an die Datenerhebung im Projekt ab: Die Daten müssen möglichst unverfälscht und in natürlichen Umgebungen erhoben werden, d.h. im häuslichen Umfeld der Interagierenden und mit einer möglichst geringen Beeinflussung durch die Erhebungssituation.

Eine Beeinflussung der Erhebungssituation scheint allerdings nahezu unumgänglich, denn um sich Zugang zu ‚natürlichen' Situationen zu verschaffen und Daten aufzuzeichnen, die nicht durch die Beforschung selbst beeinflusst werden, beeinflussen ForscherInnen nahezu immer ebendiese ‚natürliche' Situation (vgl. Ehlich 2007). Hinzu kommt, dass die Frage, was genau unter „natürlichen Gesprächen" zu verstehen ist und wann das Postulat der Natürlichkeit als erfüllt gelten kann, in verschiedenen Ansätzen umstritten ist (vgl. Gerwinski und Linz 2018, S. 109–110): So bilden Positionen wie die von Potter und Wetherell (1987), denen zufolge ausschließlich Daten verwendet werden können, die ohne Zutun und Beobachtung der ForscherInnen zustande gekommen sind, das eine Ende eines Spektrums von Natürlichkeitsauffassungen, während am anderen Ende z.B. ten Have feststellt, dass häufig nicht strikt zwischen „naturally occuring" und „experimental data" unterschieden werden kann (vgl. ten Have 1999, S. 44). So formuliert denn auch Deppermann (2008, S. 24) den Anspruch dahingehend neu, dass nicht eine ‚Natürlichkeit' gegeben sein muss, sondern dass vielmehr „das Datenmaterial und die Art seiner Erhebung und Auswertung geeignet sein müssen, die Forschungsfragen in bestmöglicher Weise zu beantworten"[1]. Wie Gerwinski und Linz (2018) am Beispiel von Pausengesprächen im Theater zeigen, kann die Beobachtungssituation durch unterschiedliche Mittel und weniger invasive Aufnahmetechniken reduziert werden. Zudem lassen sich die Beeinflussungen durch die Beobachtungssituation bei der Datenauswertung sogar produktiv wenden und etwa für metasprachliche Reflexionen nutzen.

Es sollen entsprechend im Rahmen der Pilotstudie Verfahren erprobt werden, um die Beeinflussung durch die Aufnahmesituation möglichst gering zu halten und gleichzeitig Möglichkeiten aufzuzeigen, die dennoch nie vollständig vermeidbare Beeinflussung produktiv zu nutzen.

Forschungsstand

Der sprachwissenschaftliche Forschungsstand zu Sprachassistenten in der Interaktion beschränkt sich auf einige erste Untersuchungen, die die Alltagspraktiken einzelner Familien betrachten (u.a. Porcheron et al. 2018). In diesen Studien wurde allerdings nicht systematisch das Nutzungsverhalten auf emergente und transformierte Praktiken hin untersucht, sondern sie haben immer auch die mögliche Verbesserung der „Einbettung" von Sprachassistenten in die natürliche, inkrementell entstehende soziale Interaktion im Blick. Diese Frage wollen wir umkehren und empirisch unterfüttern: Wie wird der Sprachassistent auch als solcher adressiert und wie wird dabei seine Technizität reflektiert? Wie beeinflusst diese ggf. auch den Umgang mit dem Gerät, wie wird diese sprachlich mit der sozialen Interaktion „verwoben", ohne das System als Interaktanten zu betrachten? Und nicht zuletzt auch: Wird die mit den Sprachassistenten im häuslichen Umfeld verbundene Beobachtungssituation von den Beteiligten in den erhobenen Settings explizit adressiert?

Die bereits vorliegenden Untersuchungen liefern wertvolle Anknüpfungspunkte für unsere Studie. So scheint der Nutzung eines IPA regelmäßig ein interaktiver Aushandlungsprozess darüber voranzugehen, wer das Gerät zu welchem Zweck nutzen kann und darf. Damit werden über den Gebrauch des IPA gleichsam Ordnungsstrukturen innerhalb des Haushalts (mit-)verhandelt (vgl. Porcheron et al. 2018, S. 5). Außerdem scheint es verschiedene Strategien zu geben, das Gerät aus einer laufenden Interaktion heraus zur Anwendung zu bringen: Bei einer sequenzanalytischen Betrachtung der Anwendungsfälle zeigt sich, dass es in einer Gruppe von Personen durchaus eine kommunikative Herausforderung darstellt, die Nutzung in die laufende Interaktion einzubetten. So besteht z.B. eine Strategie darin, zunächst für Stille zu sorgen, damit die Anfrage an das Voice User Interface (VUI) von diesem auch verarbeitet werden kann – die Anfrage kann aber auch so in den eigenen Redezug eingebaut werden, dass dies gar nicht erforderlich ist (Porcheron et al. 2018, S. 7–8). Die Daten von Porcheron et al. (2018) zeigen auch, dass eine Einbettung in eine laufende Konversation häufig nicht mühelos gelingt. Dies legt erneut die Vermutung nahe,

1 Dieses reformulierte Gütekriterium für die Daten wird auch als „ökologische Validität" bezeichnet, die einen produktiven Umgang mit der Beeinflussung durch die Aufnahmesituation ermöglicht, siehe dazu ausführlicher Deppermann (2008, S. 24–26) und zu einer kritischen Betrachtung des Natürlichkeitsanspruchs Gerwinski und Linz (2018, S. 114–115).

dass die VUIs eher nicht als InteraktantInnen wahrgenommen, sondern vielmehr als stimmlich zu bedienendes Gerät behandelt werden, dessen Technizität sich die Interagierenden durchaus bewusst sind. Die Geräte stellen dabei mit Krummheuer (2010, S. 317) eine „Interaktionssimulation" dar.

Überdies sind natürliche Interaktionen mit maschinellen Systemen aus sprachwissenschaftlicher Sicht v.a. für Chatbots (Lotze 2016) sowie für Embodied Conversational Agents (Pitsch 2015; Pitsch et al. 2017) untersucht worden, die erste Hinweise auf den Umgang mit artifizieller Mündlichkeit als interaktives Element in zwischenmenschlichen Interaktionen geben. Auch dabei wird konstatiert, dass die VUIs keinesfalls im Sinne eines ‚gleichwertigen' Interagierenden behandelt werden, sondern dass Anfragen bzw. Antworten an das Gerät auch Gegenstand der interaktiven Aushandlung vor dem Gerät sein können, sich also eine Art Meta-Interaktionsraum ergibt, in dem nur Menschen interagieren, um eine adäquate Formulierung bereitzustellen (Pitsch et al. 2017, S. 396). Aufschlussreiche Datenbeispiele in diesen und ähnlichen Arbeiten sind v.a. Situationen des Nicht-Verstehens, in denen Reparaturen geäußert werden. Diese Situationen zeigen häufig auch gemeinsam entwickelte Problemlösungsstrategien, wenn die Anfrage nicht erwartungsgemäß verarbeitet werden konnte. Die in den Arbeiten ausgewerteten Daten beziehen sich jedoch nicht auf häusliche IPAs, sondern überwiegend auf humanoide Roboter, die sich auch dadurch unterscheiden, dass sie nicht in gleicher Weise domestiziert und in den Alltag eingebunden sind.

Zu Intelligenten Persönlichen Assistenten selbst liegt eine Reihe Forschung aus anderen Disziplinen vor, etwa aus medienwissenschaftlicher Sicht zu medial aufgebauten Erwartungen an die Geräte (z.B. Hennig und Hauptmann 2019), zu Fragen von Privacy im Zusammenhang mit Sprachassistenten (z.B. Lau et al. 2018 und die jüngste Erscheinung der Wissenschaftlichen Dienste des Deutschen Bundestages 2019), aus Sicht der Surveillance-Studies (z.B. West 2019) oder aus Sicht der interdisziplinären Mensch-Computer-Interaktionsforschung zu anthropomorphen Zuschreibungen (siehe dazu z.B. die Interviewstudie von Krüger et al. 2018). Erwähnenswert sind auch Design-Studien, die Sprachassistenten von einer technischen Perspektive beleuchten, z.B. Dasgupta (2018) oder Pearl (2016). Gemeinsam ist diesen Studien jedoch, dass eine empirische Betrachtung von alltäglichen Praktiken im Umgang mit IPA nicht erfolgt: Die genannten Arbeiten stützen sich teilweise auf Interviews und Experimente, betrachten aber nicht die tatsächliche Nutzung durch die Mitglieder von Haushalten in ihrer sozialräumlich ‚natürlichen' Wohnumgebung. Dieses Desiderat will das Projekt schließen. Daher sind im Folgenden Herausforderungen und mögliche Lösungsansätze für die Erhebung entsprechender Daten adressiert.

Herausforderungen der Datenerhebung

Für die Datenerhebung sind auf der einen Seite Aspekte des Zugangs zu den Daten wichtig. Diese Gesichtspunkte spielen vor der eigentlichen Aufnahmesituation in den Haushalten eine Rolle, d.h. diese überhaupt für die Teilnahme an der Studie zu gewinnen und entsprechend zu instruieren. Auf der anderen Seite sind Anforderungen an die Aufnahmesituation selbst und die konkrete Vorbereitung dieser, also Themen, die während der eigentlichen Aufnahmesituation bedacht werden müssen, relevant. Diese beiden Bereiche sollen im folgenden Kapitel angesprochen und reflektiert werden. Auf die Aufbereitung und Auswertung sowie auf erste Schlussfolgerungen wird im nächsten Kapitel weiter eingegangen.

Um TeilnehmerInnen für die Studie zu gewinnen, wurden verschiedene Strategien verfolgt. Aufbauend auf den Erfahrungen aus anderen gesprächslinguistischen Projekten wurde (A) dem Ansatz einer Akquise im privaten Umfeld der ForscherInnen nachgegangen, (B) wurde im Kontext von universitären Seminaren zur Datenspende aufgerufen.

Erhebung im privaten Umfeld der ForscherInnen

Für den Ansatz der Akquise im privaten Umfeld konnte auf Erfahrungswissen zurückgegriffen werden, das z.B. in dem von der DFG geförderten Projekt „Theater im Gespräch. Sprachliche Kunstaneignungspraktiken in der Theaterpause" gewonnen wurde (vgl. Besthorn et al. 2018, S. 75; Schlinkmann und Hesse 2018).[2] Im Rahmen der Identifizierung von Haushalten mit einem häuslichen Sprachassistenzsystem konnten durch Nachfragen bei familiären oder freundschaftlichen Zusammenkünften schnell entsprechende Personen gefunden werden, die prinzipiell bereit waren, an der Studie teilzunehmen. Der Vorteil gegenüber einer Kaltakquise besteht bei diesem Vorgehen darin, dass hier bereits ein Vertrauensverhältnis zwischen den Forschenden und den Beforschten besteht und eine leichtere Zugänglichkeit ermöglicht wird.

In weiteren Anbahnungsgesprächen mit potenziellen StudienteilnehmerInnen zeigte sich jedoch, dass die Bereitschaft zur Aufnahme im häuslichen Umfeld eher gering ist, auch im Vergleich zur Be-

2 Im Rahmen dieses Projekts wurde bereits auf Projekterfahrungen von früheren Projekten zurückgegriffen, z.B. auf das von Werner Holly durchgeführte Projekt „Sprechen über Fernsehen" (1993).

reitschaft, andere Datentypen bereitzustellen. Zwei Schwierigkeiten traten immer wieder auf: Erstens waren einige, auch den ForscherInnen sehr vertraute Personen, nicht oder nur widerwillig bereit, ihre häusliche Umgebung aufzuzeichnen und zur Verfügung zu stellen. Zu groß waren die Bedenken, allzu intime Details bekanntzugeben, die nicht mit den IPA-Interaktionen in Zusammenhang standen. Dieser Befund ist auch insofern bemerkenswert, als die potenziellen ProbandInnen zwar in eine Auswertung der Daten durch die Betreiberfirmen (z.B. Amazon) eingewilligt hatten, in der Verwertung der Daten durch persönlich bekannte ForscherInnen hingegen ein Problem sahen. Dieser scheinbare Widerspruch lohnt einer weiteren Betrachtung im Rahmen des Projekts; zunächst kann aber vermutet werden, dass gerade die persönliche Beziehung zwischen Forschenden und Beforschten die angefragten Daten zu sensiblen Daten macht.

Dies verweist mit Blick auf das Natürlichkeitspostulat auch darauf, dass die Aufnahmesituationen zunächst gemeinsam zwischen ForscherInnen und TeilnehmerInnen hergestellt werden müssen: Es müssen technische Vorkehrungen getroffen werden, um eine reibungslose Aufnahme zu gewährleisten; die ForscherInnen und mehrere TeilnehmerInnen aus dem beforschten Haushalt müssen sich terminlich vereinbaren. Dies bedeutet für die StudienteilnehmerInnen einen nicht zu unterschätzenden Aufwand, den sie nicht immer zu leisten bereit waren; zudem konnten im Rahmen der Pilotstudie keine monetären Anreize geboten werden. Auch war der Zeitraum der Aufzeichnung stets beschränkt, die dauerhafte Audio-Aufzeichnung der Wohnumgebung z.B. über einen gesamten Abend oder länger lehnten alle angefragten Haushalte ab; eine Begrenzung der Aufnahmezeit auf einen bestimmten Zeitraum oder eine spezifische soziale Situation (z.B. Kochen) war allen TeilnehmerInnen sehr wichtig: Es sollten lediglich die Daten bereitgestellt werden, die auch spezifisch mit der IPA-Interaktion in Zusammenhang standen. Die Ersteinrichtung aufzeichnen zu wollen setzt voraus, dass Haushalte mit einer entsprechenden Kaufabsicht akquiriert werden müssen. Diese Kaufabsicht soll jedoch nicht durch die Forschenden initiiert, sondern intrinsisch motiviert sein, um nicht Daten von IPA-NutzerInnen auszuwerten, die ohne externes Eingreifen kein Sprachassistenzsystem hätten: dies kann erhebliche Unterschiede in Nutzungs- und Reflexionspraktik mit sich bringen, die im Rahmen des Forschungsprojekts nicht hinreichend hätten berücksichtigt werden können.

Es ist jedoch insgesamt festzuhalten, dass sich trotz der anfänglichen Zurückhaltung innerhalb weniger Wochen drei Haushalte aus dem privaten Umfeld zu einer Aufnahme von Daten und deren Auswertung im Rahmen des Forschungsprojekts bereiterklärten, davon zwei Ersteinrichtungssituationen. Dies deutet an, dass mit verbesserten Anreizen sowie einer technischen Lösung auch für die längerfristige Aufnahme von Daten in den oben erwähnten Haushalten ein zweckmäßiges Korpus aufgebaut werden kann.

Erhebung im Seminarkontext

Eine weitere Möglichkeit der Erhebung von Daten ergab sich im Rahmen eines Seminars zum Thema „Technisierte Interaktion“, das von einer am Projekt beteiligten Mitarbeiterin angeboten wurde. Die SeminarteilnehmerInnen sollten eigenständig Nutzungssituationen von VUIs (Siri oder Alexa) in ihrem privaten Umfeld aufnehmen. In einem Methoden-Block wurden sie hinsichtlich der Aufnahmemöglichkeiten (Audio- oder Videoaufzeichnungen), der mit einer Datenaufnahme verbundenen Einverständniserklärung und den für die Analyse wichtigen Prinzipien einer gesprächsanalytischen Vorgehensweise instruiert. Das Datenkorpus, das im Seminar zusammengestellt wurde, bestand einerseits aus Aufnahmen, die die Studierenden von sich selbst machten (z.B. die Nutzung des VUI auf dem Smartphone zum Nachschlagen biologischer Fachbegriffe im Rahmen einer Lerngruppe oder zur Rezeptsuche beim gemeinsamen Kochen). Andererseits besaßen einige der Studierenden oder ihnen bekannte Personen bereits ein Amazon Echo-Gerät. Die von diesen Studierenden erhobenen Daten wurden beim Zusammentreffen mit FreundInnen oder im Rahmen von familiären Zusammenkünften (z.B. beim Besuch der Großeltern) aufgenommen. Bei der Durchführung der kleinen Studie griffen alle Studierenden auf Audioaufnahmen zurück. Einige der Studierenden sowie die Personen, die an den von diesen Studierenden erhobenen Situationen beteiligt waren, gaben ihr Einverständnis, die Daten für die Auswertung im Rahmen der Pilotstudie zu spenden. Im Seminar wurden deshalb zwei verschiedene Einverständniserklärungen genutzt: eine für die Verwendung der Transkripte und gegebenenfalls Audiodateien innerhalb des Seminars und eine weitere für die Datenspende (in Form von Transkripten und ggf. auch Audioaufnahmen) im Rahmen der Pilotstudie für das Projekt B06.

Die Untersuchung im Rahmen des Seminars zeigte, dass die Studierenden vor unterschiedlichen Herausforderungen standen: Zum Teil hatten sie Schwierigkeiten, Familienangehörige oder Bekannte bzw. FreundInnen davon zu überzeugen, sich auf eine Aufnahme einzulassen (s. auch Abschnitt „Erhebung im privaten Umfeld der Forscherinnen“). Um diese Schwierigkeiten zu umgehen, nahmen sie in den meisten Fällen ihre eigenen Interaktionen auf. Als weitere Erkenntnis ging aus der im Seminar durchgeführten Untersuchung hervor, dass nur wenige der

Studierenden und der von ihnen aufgenommenen Personen bereit waren, die Daten in Transkriptform für eine Analyse bzw. Nutzung im Rahmen der Pilotstudie zugänglich zu machen. Auch zeigte sich, dass es sich weniger um aufgenommene Situationen handelte, wie sie üblicherweise im Alltag der Beteiligten stattfinden, sondern Settings geschaffen und Daten elizitiert wurden. Diese Erkenntnis unterstreicht nochmals die bereits vorangehend angesprochene Relevanz einer möglichst präzisen Instruktion bzw. Information der StudienteilnehmerInnen hinsichtlich der Zwecke der Aufnahme und der Ziele des Projekts. Trotz des quasi-experimentellen Charakters der im Seminarkontext erhobenen Daten wurde deutlich, dass auch in diesen Aufnahmen u.a. die folgenden, bereits in Studien zur Mensch-Maschine-Interaktion mit Fokus auf IPA beschriebenen Phänomene (s. Abschnitt „Forschungsstand“) auftraten:

Das multimodale Interface mancher IPAs (visuell und auditiv) kann hinsichtlich der von den NutzerInnen erwarteten Unterstützung hinderlich sein, je nach Art und Weise der Aktivität, in die der IPA eingebettet ist, z.B. wenn beim Kochen die Hände nicht frei sind, um das Touch-Display des Smartphones zu bedienen, die Antwort auf die vom Nutzer gesprochen-sprachlich realisierte Frage an das Gerät aber als schriftsprachliche Visualisierung auf dem Display ausgegeben wird (z.B. „kann gerade nicht gucken meine hände sind dreckig.“ Daraufhin wird eine Frage nach einem Bolognese-Rezept an das Assistenzsystem gerichtet, der die Sprecherreaktion „der zeigt mir nur websites.“[3] folgt).

Des Weiteren sind im Zusammenhang mit dem Zweck der Nutzung nutzergruppenspezifische Einstellungen hinsichtlich des Geräts in den Ausschnitten sichtbar, die mitunter mit Anthropomorphisierungen des Geräts (s. auch Abschnitt „Mögliche Phänomenbereiche und Analysekategorien“) einhergehen: So zeigt sich einerseits Bewunderung, was das Gerät kann (in den im Seminar erhobenen Beispielen v.a. bei ErstnutzerInnen), andererseits aber auch Frustration und Ablehnung, weil die Technik nicht wie erwartet funktioniert (z.B. die Äußerung „MEIne güte ist die DUMM“[4] nach dem mehrmaligen Versuch, das Gerät nach einer an es gerichteten Frage zum Antworten zu bringen).

Im folgenden Beispiel (1) führt Nutzerin Christa (C) ihrer Bekannten Fabienne (F) das in ihrem Besitz befindliche IPA Amazon Echo (A) vor.

Beispiel (1)[5] : Erstnutzung/Vorführung Alexa

```
019  C:  ich weiß nich wenn ich sie
         jetz fragen würde: (-)
020      ehm was kann man sie denn mal
         googlen lassen?
021  p:  (1.5)
022  C:  hm: (0.7) alexa:? (0.9) was ist
         aquaplaning;
023  p:  (2.1)
024  A:  aquaplaning bezeichnet das
         aufschwimmen des reifens
025      auf dem wasserfilm einer nassen
         fahrbahn.
026  p:  (0.7)
027  F:  sie is so klug
028  p:  (0.2)
029  C:  ja (-) [aber ansch]einend kann
         sie doch g[oo]glen
030  F:         [ ((lacht))]
                [ja] °h
```

In Zeile 019 manifestiert sich bei Christa zunächst Unsicherheit, welche Frage an das Gerät gerichtet werden soll, gefolgt von ihrem Metakommentar, was man „sie denn mal googlen lassen“ kann, der zugleich einen Hinweis darauf gibt, dass Christa der Zugriff des VUI auf die Suchmaschine Google und damit externe Datenquellen bewusst ist. Nach einer an den Sprachassistenten gerichteten Informationsfrage von Christa gibt dieser die entsprechende Antwort aus. Dies führt Christas Bekannte Fabienne zur Prädikation „sie is so klug“ (Z. 027), in der Fabienne die von Christa bereits vorangehend realisierte weibliche Genderzuschreibung „sie“ aufgreift, um auf das Gerät zu referieren.

In Beispiel (2) erprobt der 73-jährige Arthur (AR) erstmals das Assistenzsystem auf dem Smartphone (SI). In diesem Ausschnitt ist außerdem seine Ehefrau Anette (AN) anwesend.

Beispiel (2): Erstnutzung/Vorführung Siri

```
008  AR:  SIri (1.0) wie ist das
          wetter in rothenburg (.) ob.
009       der.
010       tauber.
```

3 Hier wird zu Illustrationszwecken ausschließlich die betreffende Stelle zitiert, da das Transkript nicht zur Verwendung freigegeben wurde. Die konsequente Kleinschreibung ist dadurch begründet, dass der Ausschnitt als Minimaltranskript verschriftet wurde.

4 Hier wird zu Illustrationszwecken ausschließlich die betreffende Stelle zitiert, da das Transkript nicht zur Verwendung freigegeben wurde.

5 Bei den im Abschnitt „Erhebung im Seminarkontext“ präsentierten Ausschnitten handelt es sich um Datenbeispiele, die Studierende im Seminar „Technisierte Interaktion“ im Sommersemester 2019 an der Universität Siegen erhoben haben. Sowohl die an der Erhebung der jeweiligen Beispiele beteiligten Studierenden sowie die weiteren an den erhobenen Situationen beteiligten Personen gaben ihr Einverständnis zur Nutzung der anonymisierten Ausschnitte in Transkriptform im Rahmen der Pilotstudie im Projekt B06.

```
011  SI:  so wird das wetter in
          rothenburg ob der tauber
          heute.
012  AR:  a:ch guck dir DAS an.
013       ACHTundzwanzig grad.
014       ZACK hat sie_s hier.
015       das ist ja sa:genhaft.
016       sa:genhaft guck dir DAS an.
017  AN:  hm
018  AR:  DIE ist ja clever (.) das
          mädchen.
019  AN:  ja echt [clever.]
020  AR:          [hm     ]
```

Arthur stellt dem VUI eine Frage nach dem Wetter in Rothenburg ob der Tauber. Siri referiert mit dem deiktischen Verweis „so“ (Z. 011) auf vermutlich schriftliche Informationen auf dem Display des Smartphones, das Arthur in der Hand hält. Seine Reaktion „a:ch guck dir DAS an. ACHTundzwanzig grad. ZACK hat sie_s hier.“ (Z. 012-014) weisen darauf hin, dass Arthur die von ihm eingeforderte Information auf dem Display angezeigt wird, auf das er mit dem deiktischen Verweis „hier“ (Z. 014) referiert. Es folgt eine explizit positive Bewertung Arthurs in Form der Deklaration „das ist ja sa:genhaft.“ (Z. 015). In Zeile 016 wiederholt er das positiv evaluierende Adjektiv „sa:genhaft“ in derselben prosodischen Konturierung wie in seiner vorangegangenen Äußerung und schließt daran die formelhafte Phrase „guck dir DAS an“ (Z. 016) an. Anettes Reaktion in Zeile 17 lässt darauf schließen, dass sie Arthurs Äußerung möglicherweise den Status einer an sie adressierten Aufforderung zuschreibt. In Zeile 18 kommt es zu einer Anthropomorphisierung des Geräts als „mädchen“, dem Arthur zudem durch eine weitere Prädikation menschliche Eigenschaften zuschreibt („DIE ist ja clever“).

In der Hauptstudie des Projekts wird – über die bei der Analyse der beiden Beispiele in den Blick genommenen sprachlich-interaktionalen Praktiken hinaus – der Fokus zudem darauf gerichtet, inwiefern die Einbettung der IPAs in die menschliche Interaktion in weitreichendere soziokulturelle Praktiken eingebettet ist (s. dazu auch die Einführung in diesen Beitrag). Denkbar wären zum Beispiel Praktiken der Weitergabe und Aushandlung von Wissen im Gespräch (in deren Rahmen eine Informationsfrage an das Gerät gerichtet wird) sowie die (gemeinsame) Planung von Alltagsaktivitäten, in die eine Frage nach dem Wetter an einem bestimmten Ort eingebettet sein könnte.

Aufnahmesituation

Einige der Herausforderungen der Datenerhebung ergaben sich retrospektiv auch während der Aufnahmesituation. So waren zwar mündlich einige Instruktionen an die StudienteilnehmerInnen übermittelt worden, doch waren diese offenbar nicht spezifisch genug. So fertigte ein Haushalt für den Situationstyp der Ersteinrichtung ein Video an, in dem einem unsichtbaren „Zuschauer“ ähnlich wie in einem YouTube-Tutorial die Ersteinrichtung erläutert wird:

Beispiel (3): Ersteinrichtung (Alexa 2)[6]

```
001       +*(0.5)
                +*
002  b:   +schaut auf Handy------------>
          >----+
003  l:    *nimmt sich Anleitung und
          Kabel--*
004  B:   (so) (2.0)
005       dann (2.0) so
006       *die app wurde jetzt
          RUNtergelAden;*
     b:   *hält handy in die kamera----->
          >-------------*
007       (-) und em (1.0)
008  B:   jetzt *gehts WEIter hier,*
     b:          *tippt auf Handy--*
009  B:   *mit dem (.) EINrichten der App
          als Erstes;  *
     l:   *versucht kabel an alexa
          anzuschließen*
010       +*(3.0)           +
011  b:   +schaut auf Handy+
012  B:   *jA ich möchte ein geRÄT
          einrIchten;*
     b:   *tippt auf Handy-------------->
          >----------*
```

6 Die Ausschnitte wurden entsprechend der Notationskonventionen des Gesprächsanalytischen Transkriptionssystems 2 – GAT 2 (Selting et al. 2009) sowie gemäß der Konventionen für multimodale Transkripte (Mondada 2014) transkribiert. Eine Übersicht über die Konventionen findet sich im Anhang. Für Arbeiten an den Transkripten danken wir Viviane Börner, Franziska Niersberger-Gueye und Franziska Petri.

In dem vorliegenden Datenbeispiel richten Beate, Lukas und Hendrik den neu angeschafften Sprachassistenten ein.[7]

Beate wendet sich dabei zu Beginn der Einrichtungssituation nicht an ihre anderen Haushaltsmitglieder, sondern zeigt ihr Handydisplay, um den Prozess der Einrichtung für die antizipierten ZuschauerInnen sichtbar zu machen:

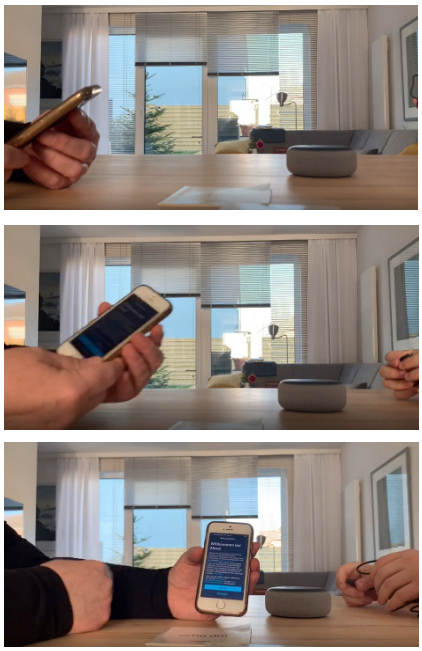

Abb. 1: Ausschnitt aus „Ersteinrichtung (Alexa 2)", Standbilder bei 00:05, 00:10 und 00:12 (Segmente 001 bis 013).

Dieser Ausschnitt zeigt, dass Beate ihr Handeln nicht (nur) gegenüber den anderen Haushaltsmitgliedern „accountable" macht (Garfinkel 1967, S. 36), sondern auch gegenüber der Kamera bzw. gegenüber nicht-anwesenden InteraktantInnen. Hier zeigen sich Parallelen zu multimodal realisierter Parainteraktion etwa auf YouTube, in denen ebenfalls die Ressourcen der Objektpräsentation im Verhältnis zur Kamera von den Agierenden genutzt werden, die somit die „Illusion erzeugen, sie seien mit den Userinnen face-to-face in einem gemeinsamen Raum (und vice versa)" (Böckmann et al. 2019). Im weiteren Verlauf des hier gezeigten Ersteinrichtungsgesprächs wird deutlich, dass die Orientierung auf die Kamera abnimmt und sich Beate auch an Lukas und Hendrik wendet; immer wieder nimmt sie jedoch starken Bezug auf die und erklärt ihr Handeln gegenüber den nicht-anwesenden Dritten. Dabei zeigt sich bei Beate auch auf der prosodischen Ebene eine deutlich veränderte stimmliche Modulation (erhöhte Lautstärke, klarere Akzentuierung und kürzere Intonationsphrasen).

In dem hier gezeigten Ausschnitt wird zudem deutlich sichtbar, dass die Interagierenden den Bildausschnitt so gewählt haben, dass zwar das Geschehen auf dem Smartphone-Display von Beate sowie auf dem Sprachassistenten selbst nachvollziehbar wird (einschließlich der Verkabelung), gleichzeitig aber die Gesichter und Körper der Interagierenden ausgeblendet werden. Lediglich die Hände sind noch partiell im Bild. Damit geht das Potenzial einer multimodalen Analyse z.B. der Gestik und Mimik der zwischenmenschlichen Interaktion vor dem Gerät verloren, die insbesondere aufschlussreich wäre, da der auditive Sprachkanal bereits durch die Interaktion mit dem Sprachassistenten ‚belegt' ist und insofern angenommen werden kann, dass ein Großteil der Aushandlungs- und Bewertungsaktivitäten der Interagierenden über einen non-verbalen Sprachkanal vermittelt wird. Eine für die Datenerhebung im Rahmen der Hauptstudie wichtige Erkenntis, die aus diesem Beispiel abgeleitet werden kann, ist zudem, dass das Geschehen auf dem Bildschirm des Smartphones unbedingt mit aufgezeichnet werden sollte.[8]

Es kann konstatiert werden, dass sich die Instruktionen, die den Haushaltsmitgliedern vor der Datenerhebung gegeben worden waren, als nicht spezifisch genug erwiesen haben, um den avisierten Datentyp tatsächlich vollumfänglich aufnehmen zu können. Dies muss einerseits in Hinweisen für die Wahl des Bildausschnitts reflektiert werden, andererseits muss verdeutlicht werden, dass die GesprächsteilnehmerInnen den Prozess keineswegs für unbeteiligte Dritte erklären und „nachvollziehbar" machen sollen, sondern die Ersteinrichtung möglichst unbeeinflusst und ohne Orientierung auf die Kamera stattfinden soll. Bei der Formulierung dieser Instruktionen ist wiederum darauf zu achten, dass eine zu starke Anleitung der Aufnahme durch die ForscherInnen zu vermeiden ist, um das Natürlichkeitspostulat nicht durch entsprechende Auflagen zu unterminieren. Insgesamt empfiehlt sich die Anwesenheit der Forschenden, die die technische Seite betreuen und einen geeigneten Bildausschnitt auswählen können.

7 Beate ist 55 Jahre alt, Lukas und Hendrik sind ihre Söhne, sie sind 18 und 20 Jahre alt. Alle in diesem Aufsatz gezeigten Datenbeispiele und Angaben über die SprecherInnen wurden vollständig pseudonymisiert und verfremdet, um die Wiedererkennung realer Personen so weit wie möglich zu erschweren.

8 Für eine solche Bildschirmaufnahme wurden inzwischen sowohl für Android- als auch für iOS-Geräte entsprechende Apps entwickelt, die als Freeware erhältlich und mit wenig Aufwand zu installieren sind.

Maßnahmen

Aus den Erfahrungen im Projekt konnten vielversprechende Lösungsansätze entwickelt werden, die die Datenerhebung vereinfachen sollen:

(1) Für die Aufnahmen sollen in Anwendung des Schneeballprinzips und unter vermehrter Durchführung von öffentlichen Aufrufen Haushalte gewonnen werden, die den ForscherInnen zwar nicht gänzlich unbekannt sein müssen, aber nicht in gleicher Weise vertraut sind. Der Suchradius soll entsprechend erweitert werden.

(2) Die Aufnahmen im häuslichen Umfeld außerhalb der Ersteinrichtung sollen mithilfe eines Conditional Voice Recorders (CVR) stattfinden, der bereits in Arbeiten von Porcheron et al. (2018) zum Einsatz kommt.[9] Durch diese Hardware-Lösung ist es möglich, nur dann Audio-Signale aufzuzeichnen, wenn ein bestimmtes Aktivierungswort genannt wird. Wird dieses an das Aktivierungswort des Sprachassistenzsystems angeglichen (z.B. „Alexa" oder „Siri"), wird nur dann die Aufnahme gestartet, wenn auch eine IPA-Interaktion erfolgt. Um zusätzlich nachverfolgen zu können, wie sich die IPA-Nutzung anbahnt, kann der CVR so programmiert werden, dass eine angemessene Zeit vor der Nennung des Aktivierungsworts ebenfalls gespeichert wird. Dadurch werden relevante Daten gesammelt, gleichzeitig wird aber die dauerhafte Aufnahme des häuslichen Umfelds vermieden.

(3) In Erhebungssituationen für die audiovisuelle Aufzeichnung der Ersteinrichtung ist die Vorbereitung des Settings durch die Forschenden sinnvoll. Diese können im Vorfeld der eigentlichen Aufnahme die Kameras einrichten, Bildausschnitte wählen und Fehler vermeiden, die zu einer Unbrauchbarkeit der Daten führen würden.

Wie sich an Beispiel (3) zeigt, ist die zum Sprachassistenten gehörige Smartphone-Applikation für die Ersteinrichtung von gravierender Bedeutung. Die Interagierenden orientieren sich häufig eher auf die App denn als auf den Sprachassistenten selbst, wie auch am folgenden Beispiel deutlich wird:

Beispiel (4): Ersteinrichtung Alexa 2

```
062  L:  in welchem RAUM befindet sich-
         (.)
063      °hh echo DOT;
064      (-) im (.) im WOHNzimmer.
065      (2.0)
066  L:  ((sucht in der App die
         passende Kategorie))
         WOHNbereich.
067      (9.0)
068  L:  ((nimmt das Handy und hält es
         Beate hin))
069  B:  ((nimmt das Handy und liest))
070      alexa verwendet diesen NAmen um
         sie besser? (2.0)
071  A:  NACHname bedeutet im
         unterschied zum vOrnamen-
072      der NACHgestellte name;
073      (.) der [faMIlienname-]
074  L:          [hh°           ]
075  A:  der im dEUtschen geSCHLECHtlich
         nicht unterschieden wird.
076      (-) übrigens?
077      (.) wenn du mir deine STIMme
         beibringst kann ich deine alexa
         erfahrung persönlicher
         gestAlten.
078      möchtest du dir dafür jetzt
         einen moMENT zeit nehmen?
079  B:  !NEIN,!
080      (19.0)
081  B:  ((gibt das Handy einem weiteren
         Sprecher))
082  L:  <<flüsternd> du hast jetzt bei
         (    ) VORnamen eingegeben;>
083      (2.0)
084  u:  ((unterdrücktes, sehr leises
         Lachen))
085  B:  so dann geb_ich jetzt den
         VORnamen ein,
```

Das Handy wird mehrfach hin- und hergereicht (Z. 068f., Z. 079). Dies und die langen Sprechpausen (Z. 067, Z. 078) sowie darüber hinaus die geflüsterten und bemüht leisen Sequenzen (Z. 080 und 082) deuten darauf hin, dass hier Kommunikation stattfindet, die sich auf die Smartphone-App orientiert, die bewusst nicht mündlich vorgetragen wird. Dies kann darin begründet liegen, dass der mündliche Kommunikationskanal bereits durch die Kommunikation mit dem Sprachassistenten belegt ist und für die Kommunikation zwischen den GesprächsteilnehmerInnen ein anderer Kanal gefunden werden muss; denkbar wäre allerdings auch, dass hier ein (virtueller) „Kommunikationsraum" geschaffen wird, der bewusst die Aufzeichnung für das Forschungsprojekt ausschließt. Die Relevanz multimodaler Analysen und AV-Aufnahmen gerade für die Ersteinrichtung findet sich hier bestätigt.

9 Für die Bereitstellung einer Dokumentation über die Entwicklung des CVR sowie weiterführende Hinweise danken wir Martin Porcheron und Stuart Reeves vom Mixed Reality Laboratory der University of Nottingham. Für die Unterstützung bei der Konfiguration des Conditional Voice Recorders und die Adaption für unser Projekt danken wir den MitarbeiterInnen im Fab Lab der Universität Siegen, insbesondere Fabian Vitt.

Mögliche Phänomenbereiche und Analysekategorien

Anhand der folgenden Beispiele soll ein erster Einblick in mögliche Phänomenbereiche gegeben werden, die für die sprachwissenschaftliche Untersuchung relevant sein könnten. Aus diesen Anhaltspunkten ergeben sich im Sinne eines datengeleiteten Vorgehens spätere Analysekategorien, zu denen wir bereits im Rahmen dieser Pilotstudie einige erste Erkenntnisse sammeln konnten.

Zunächst scheint eines der in der Mensch-Maschine-Forschung bereits herausgestellten Phänomene auch in den im Rahmen der Pilotstudie erhobenen Daten relevant: Der Umgang mit Koordinationsschwierigkeiten zwischen Mensch und Maschine. Die folgenden Beispiele (5) bis (10) stammen aus einer Interaktion der vier Geschwister Jan, Josefine, Markus und Hendrik.[10]

Beispiel (5): Reh

```
001  J:  [unGLAUblich, hh°     ]
002  A:  [((ausklingender Ton))]
003      ein jUnges von welchEm tIEr
         nennt man KITZ?
004      (0.6)
005  M:  REH,
006      (1.4)
007  A:  entSCHULDigung (-) ich hAbe
         dich nicht verstAnden.
008      (0.4)
009  M:  RE[H, ]
010  A:    [ein] jUnges von welchEm tIEr
         nennt man KITZ?
011      (0.7)
012  M:  !REH!;
```

An der Oberfläche zeigt sich hier, dass der Sprachassistent Schwierigkeiten hat, das einsilbige Wort „Reh“ zu verarbeiten und entsprechend die erneute Eingabe anfordert. Das anschließende Turn-taking verläuft jedoch nicht reibungslos: Die „Turnübernahme“ durch den menschlichen Sprecher erfolgt zu früh, denn der Sprachassistent wiederholt die zu einem Quiz-Spiel gehörende Frage noch einmal. Dies hatte Markus offensichtlich nicht antizipiert und seinerseits den Turn („REH“, Z. 009) „zu früh“ produziert. Es kommt zu einer Überlappung und da sich der Sprachassistent nicht im Aufnahme-, sondern im Abspielmodus befindet und die Äußerung von Markus nicht parallel verarbeitet, ist zu einer im Sinne des Spielverlaufs erfolgreichen Bedienung des Sprachassistenten eine erneute Eingabe erforderlich, die sich denn auch durch erhöhte Lautstärke und prosodische Akzentuierung auszeichnet. Auffällig ist, dass in diesen Vorgang von den anderen Anwesenden nicht eingegriffen wird und auch in den folgenden Sequenzen keine Bewertung erfolgt. Dies kann als ein Hinweis auf das häufige Auftreten entsprechender Koordinationsprobleme gelesen werden.

Im Fokus soll aber neben der ‚reinen‘ Untersuchung der Koordinationsprobleme selbst außerdem die Betrachtung der interaktiven Bearbeitung solcher Koordinatonsprobleme stehen, wie sich etwa im folgenden Gesprächsausschnitt zeigt:

Beispiel (6): USA

```
007  A:  in wElchem LAND ist der
         indiAnapolis speedway
         behEImatet?
008      (1.3)
009  M:  u es A;
010  A:  ((zweiteiliger Ton))
011  U:  [hä?  ]
012  A:  [(rom)]- ist FALSCH.
013      (0.5)
014  A:  es wAr vereinigte STAATen.
015      (0.3)
016  J:  es war FALSCH maArkus.
```

Es kommt hier zu einem Koordinationsproblem, weil das Gerät „USA“ als Antwort nicht erkennt und nur die in diesem Kontext gleichbedeutende Antwort „Vereinigte Staaten“ als zulässig erachtet. Josefine betont hier durch eine Wiederaufnahme der syntaktischen Struktur sowie durch eine zusätzliche personale Adressierung an den Spieler den offensichtlichen Funktionsfehler des Geräts. Sich daran anschließende Fragen sind: Welche anderen Formen der interaktiven Bearbeitung von solchen oder ähnlichen Koordinationsproblemen und „Missverständnissen“ können identifiziert werden? Welche sind bereits musterhaft im Gebrauch verankert und können als sedimentierte Praktiken beschrieben werden? Wo greifen die NutzerInnen dabei auf ein bestehendes Repertoire an Praktiken zurück, das bereits in anderen kommunikativen Konstellationen zum Einsatz kommt, wo bilden sich ggf. auch neue Praktiken der Bewältigung von Koordinationsproblemen? Welche Agency wird dabei dem Sprachassistenzsystem zugeschrieben, wie werden ko-präsente SprecherInnen in die Bewältigung mit einbezogen?

Im Hinblick auf die theoretische Konzeptualisierung muss zudem hinterfragt werden, inwiefern Termini wie „Verstehen“ und „Missverstehen“ hier überhaupt angebracht sind, wo es sich doch um einen maschinellen Sprachverarbeiter handelt, der nicht mental versteht, aber im Sinne Deppermanns (2013, S. 1) in „sprachlich-kommunikativ[en] Verfahren der Dokumentation von Verstehen“ durchaus zum Aus-

10 Jan ist 27 Jahre alt, Josefine 21, Markus 20 und Hendrik 18. Sie kochen gemeinsam bei Josefine.

druck bringt: Auch ein Sprachassistent zeigt an, ob die Äußerung hinreichend spezifisch war, um das gewünschte Ergebnis zu liefern, mithin, ob er sie „verstanden" hat. Diese Diskussion eines weitreichenden Konzepts wird im Rahmen des Projekts ebenfalls zu führen sein.

Darüber hinaus scheint ein sprachwissenschaftlich relevanter Untersuchungsbereich die regelmäßige Anthropomorphisierung des Geräts zu sein. Wiederholt treten in der bisher vergleichsweise kleinen Datenbasis entsprechende Zuschreibungen auf:

Beispiel (7): Miauen

```
001  A:  ((miaut vielstimmig))
         [((miaut vielstimmig))]
002  U:  [((leises Lachen))    ]
003  A:  ((miaut vielstimmig))
         [((miaut vielstimmig))]
004  J:  [°hh so GEIl-         ]
005  A:  [((miaut vielstimmig))]
006  M:  [mh AlexA?            ]
007  A:  ((miaut vielstimmig))
         [((miaut vielstimmig))]
008  M:  [aLEXa?               ]
009  J:  die miAUT grAde.
010      (0.2)
011  M:  apPLAUS?
012      (0.9)
013  A:  ((Händeklatschen)) ((Trommeln))
         ((Händeklatschen))
```

Markus versucht eine Anfrage an den Sprachassistenten zu richten, indem er diesen wiederholt adressiert (Z. 006 und 008). Die an Markus gerichtete Erklärung von Josefine, warum der Befehl nicht verarbeitet wird („die miAUT grade", Z. 009) baut nicht auf einer Erklärung der Gerätefunktionen auf, sondern schreibt dem Gerät menschliche Eigenschaften zu. Es wird dabei auch grammatisch zum Subjekt und zudem mit einem weiblichen Genus belegt. Bevor darauf näher eingegangen wird, soll zunächst eine etwas anders gelagerte Anthropomorphisierung im folgenden Beispiel präsentiert werden:

Beispiel (8): Voll eingeschnappt, die Gute

```
001  J:  [AUa aua aua.]
002  M:  aLEXa (.) stOp;
003  A:  entSCHULdigung (-) ich hAbe
         dich nicht ver[stAnden.]
004  J:                [SORry;  ]
005  M:  aLEXa;
         !STOP,!
006      (0.8)
007      EN[de.]
008  J:            [o::ch.      ]
009  A:    [ du] hast [STOP gesagt.]
010  A:  mÖchtest du das
         [SPIEL beEnden?          ]
011  M:  [ja es hat ja KEInen sinn.]
012  H:  alexa sEI
         [STILL;                  ]
013      [((klirrendes Scheppern))]
014       (0.5)
015  M:  JA;
016  J:  och: (.) d[ie !A:R!me.]
017  A:            [oKAY (--)  ] dAnke
         fürs spIElen.
018      (0.2)
019  J:  okA[::Y?   ]
020  A:     [wenn du] irgendein feedback
         für (.)
021      un[ser TEAM hast (.) lasse es
         uns]
022  J:    [voll EINgeschnappt die gUte;
           ]
023  A:  [(   )    ] WISSen unter (ei) ät
         lAbworks punkt I O.
024  M:  h° HE he;
```

Diese Form der Vermenschlichung bezieht sich nicht auf eine menschliche Eigenschaft, die dem Sprachassistenten zugeschrieben wird, sondern eine menschliche Emotion. Diese Zuschreibung einer Emotionalität erfolgt in diesem Ausschnitt sogar zweifach: zunächst als erst Markus und anschließend Hendrik einen Befehl an den Sprachassistenten äußern, mit dem die Beendigung des aktuellen Programms hervorgerufen werden soll (Z. 005–007, Z. 012). Plötzliche Unterbrechungen in Gesprächen werden von Betroffenen als störend und aggressiv empfunden (vgl. Linke et al. 1996, S. 267) und rufen auch bei Josefine eine empathische Reaktion hervor, zunächst nur durch die mit fallender Tonhöhenbewegung geäußerte Interjektion „o::ch" (Z. 008), kurz darauf jedoch auch noch einmal expliziter („die !A:R!me", Z. 016). Neben der Verwendung der Interjektion „och", die per se schon typisch für den Ausdruck von Emotionalität in der Interaktion ist, äußert Josefine damit zusätzlich eine Mitleidsbekundung und zeigt ihre Offenheit für eine Perspektivenübernahme. Sie präsentiert damit nach Fiehler (2002, S. 96–97) typische Anzeichen für Emotionalität in der Interaktion.[11] Durch die starke Akzentuierung, die Dehnung und die fallende Tonhöhenbewegung (Z. 016) wird der Emotionalität auch prosodisch Ausdruck verliehen. Mit „oKAY dAnke fürs spIElen." (Z. 017) reagiert der Sprachassistent schließlich auf den Sprachbefehl von Markus und beendet das Programm. Josefine nimmt anschließend die Äußerung

11 Zu Mitleidsbekundungen und Empathie in der Interaktion siehe außerdem Kupetz (2014, 2015).

des Sprachassistenten auf („oKA::Y“, Z. 019). Sie schreibt hier nun dieser Äußerung eine Emotionalität zu, indem sie die „Bewertung“, die der Sprachassistent hier mit Blick auf den Sprachbefehl vollzieht, einer Interpretation unterzieht und sie als „voll EINgeschnappt die gUte“ (Z. 022) anthropomorphisiert. Dabei ist im Aufbau der Äußerung insbesondere die Wiederaufnahme der Antwortpartikel „okay“ relevant, die durch ihre prosodische Akzentuierung (Dehnung, stark steigende Tonhöhenbewegung, längere Pause vor der nächsten Intonationsphrase) die zusätzliche Aufladung des „Erkenntnisprozessmarkers“ (Imo 2009) zu einem emotionalen Ausdruck vorbereitet. Sie überzeichnet damit ursprüngliche Merkmale, sodass sie eine neue Interpretation ermöglichen. Markus ratifiziert diese Zuschreibung als witzig, indem er Lachpartikeln äußert (Z. 024).

Die gezeigten Beispiele werfen verschiedene Fragen auf: In welchen Situationen und Konstellationen kommt es zu einer mehr oder weniger humorvoll aufgeladenen Zuschreibung menschlicher Eigenschaften, Motivationen, Intentionen oder Emotionen (vgl. Epley et al. 2007)?[12] Wie werden diese sequenziell eingebettet und kommunikativ funktionalisiert? Wie werden sie sprachlich gestaltet? Inwieweit wird dem Sprachassistenten eine Agens-Rolle zugeschrieben und von den Interagierenden selbst als (teil-) autonomer Handlungsakteur konzeptualisiert und auf welche Weise reflektieren die Interagierenden diese Zuschreibungen? Welche Rolle spielen Gender-Kategorien dabei und wie wird der Zusammenhang zwischen der (im Regelfall) weiblichen Stimme des Sprachassistenten und der Reflexion über die Rolle des Geräts im Sinne eines ‚doing gender‘ (Garfinkel 1967) als „fortlaufende Bewerkstelligung“ (West und Zimmermann 1989) interaktional konstruiert und sprachlich hervorgebracht?[13] Diese Punkte werden in der Hauptstudie zu adressieren sein.

Nicht zuletzt soll auch die „accountability“ für die Nutzung des Sprachassistenten in den Blick gerückt werden: So reagieren die Interagierenden z.B. auf die Nutzung des Sprachassistenten durch andere Interagierende wie im folgenden Beispiel:

Beispiel (9): Fordere mich heraus

```
001  M:  aLEXa?
002      (0.8)
003  M:  alEXA?
004      (0.8)
005      fOrdere mich herAUS.
006  J:  och NE::.
007      (0.5)
008  M:  DOCH-
009      (0.2)
010  A:  hIEr ist der skill TRIvia herO
         (-)
         on [labwOrks punkt I o de e.   ]
011  M:     [is echt COOL.             ]
012  A:  ((anklingender Ton))
         ((Melodie))
```

Während in diesem Ausschnitt Markus den Skill „Fordere mich heraus“ spielen bzw. vorführen möchte, zeigt Josefine durch die Kombination des change-of-state-token „ach“ (vgl. Golato 2010) mit der hier negativ gebrauchten Antwortpartikel „ne“ an, dass diese Handlung unerwünscht ist (Z. 006). Entsprechend reagiert Markus auch mit „DOCH“ und macht anschließend „accountable“, seine Nutzung fortzusetzen („is echt COOL“, Z. 012). Dabei spricht er parallel zur Äußerung des Sprachassistenten, den er nicht in dieses Geschehen einbezieht, aber dessen Eigenschaften (bzw. die Eigenschaften des Skills) er zur Aushandlung seiner Nutzung heranzieht.

Hinweise auf Praktiken der Interaktion vor den Sprachassistenzsystemen, die zwar unmittelbar in Zusammenhang mit diesen stehen, sich aber nicht an diese richten, liefert auch das folgende Beispiel, in dem Markus einen Skill vorführen will, dieser allerdings mehrfach nicht funktioniert wie geplant:

Beispiel (10): „Sag meiner Oma“

```
001  M:  wir haben ne neue funktiON
         entdeckt (-) übrigens;
002      [und ZWAR,       ]
003  J:  [wusstet ihr dass] fEta so
         krAss SCHMILZT (.) lecker.
004      (1.2)
005  J:  [oh;]
006  M:  [oh ] (-) KÖSTlich.
007  J:  m:?
008      (1.1)
009  M:  aLEXa?
010      (0.3)
011  J:  m:[::?            ]
012  M:    [ÖFFne den skill,]
013      (0.2)
014  M:  sag MEIner Oma.
015      (1.7)
016  A:  ich kann diesen skill nicht
         [FIN]den.
017  M:  [hä?]
018      (0.4)
019  M:  [aLEXa?]
020  A:  [SKILLS] findest du-
021  J:  halt du musst das ein[fach
```

12 Für einen Überblick zu Anthropomorphisierungen in Mensch-Roboter-Interaktionen siehe auch Marquardt (2017). Speziell zu (humanoiden) Robotern und Emotionen siehe Habscheid et al. (2019).

13 Zum Konzept des doing gender aus linguistischer Perspektive siehe auch Kotthoff (2003).

```
        nur sagen-]
022  M: [sag meiner
                  ][Oma,   ]

((Auslassung, 5 Sek.))

023  A:             [DERzeit]
024     wird diese kaLENderabfrage
        nicht unterstützt.
025  J: das ist so MERKwürdig grade.
026  M: aLEXa::?
027  J: das is als ob wir uns diesen
        kOmischen befEhl AUSgedacht
        hätten.
028  U: jAha.
029  M: aLEXa?
030  M: sAg mEIner Oma,
031     (0.4)
032     dass (-) mein (-)FAHRzeugbrief
        in der blAUen schUblade ist.
033     (2.7)
034  J: <<p>DEIner?>
035  A: darauf habe ich
        lei[der KEIne an]twort.
036  M:    [aLEXa?      ]
037     (0.3)
038  M: sag deiner Oma,=
039     =meine PFANne (-) ist in der
        schUblade.
040     (0.7)
041  A: aLEXas Oma ist aktiVIERT;
042     (0.4)
043  U: ((Lachen))
044  A: möchtest du den SKILL öffnen?
045     (0.3)
046  M: !JA;!

((Auslassung der Erläuterung durch das
System, 55 Sek.))

047  M: frage deine Oma nach meinem
        FAHRzeugbrief;
048     (0.7)
049  A: aLEXas Oma ist aktiVIERT;
050  J: <<pharyngalisiert, f>!o::ch.!>
051  M: alEX[a?  ]
052  A:     [möch]test du den-
053  J: alExas oma <<f>!KLAPPT!>
        [einfach nich.]
054  M: [wO Ist       ] mEIn
        !FAHR!zEUgbrIEf.
055  U: (-) ((Lachen))
056  A: aLEXas Oma ist aktiVIERT;
057  J: och MANN-
058  M: a!LEX!a?
059  A: möchtest du den-
060  J: ach a!LE[X!a;     ]
061  M:         [ sage dei]ner Oma dass
        sie meinen FAHRzeugbrief
        vergessen soll;
062  A: ((zweiteiliger Ton))
```

Markus will eine bestimmte Funktion „vorführen“. Zunächst kann der entsprechende Skill zwei Mal nicht gefunden werden (Z. 019, 024), was bereits zu einer ersten Bewertung führt („als ob wir uns diesen kOmischen befEhl AUSgedacht hätten“, Z. 027). Auch der dritte Spracheingabebefehl wird nicht erwartungsgemäß ausgeführt (Z. 035), während der vierte Befehl (Z. 041) immerhin partiell als ‚verstanden‘ betrachtet werden kann, ohne dass der eigentliche Befehl umgesetzt wird; es wird zunächst das Programm geöffnet und anschließend erläutert. Als anschließend ein entsprechend der Programmerläuterung formulierter Sprachbefehl geäußert wird, wird auch dieser nicht wie erwartet verarbeitet, sondern zwei Mal lediglich der Programmstart ‚vermeldet‘ (Z. 049, Z. 056) und die Erläuterung begonnen, die jeweils abgebrochen wird. Die ko-präsente Sprecherin Josefine bewertet die Funktionsweise des Sprachassistenten negativ und bringt dies durch eine Interjektion (Z. 050) zum Ausdruck. Außerdem produziert sie eine ebenfalls negativ zu klassifizierende Gesamtbewertung des Skills („alExas oma !KLAPPT! einfach nich“ Z. 053). Die zusätzlich produzierte Gesprächspartikel „ach“[14] (Z. 058) in Kombination mit der personalen Referenz, die hier nicht als Signalwort für einen Sprachbefehl, sondern vielmehr als „Geste personifizierter Zuwendung“ (Günthner 2016) betrachtet werden kann, zeigt den Unmut der Sprecherin Josefine, der hier erneut in Form einer Anthropomorphisierung ausgedrückt wird. Nicht im Transkript dargestellt, aber in unmittelbarer Folge zeigen sich weitere Bewertungs- und Anthropomorphisierungshandlungen.

Beispiel (10) macht deutlich, dass Erwartungen an den Sprachassistenten sowie der Umgang mit einer nicht erwartungsgemäßen Reaktion interaktional ausgehandelt werden. Dass Anthropomorphisierungen dabei als Ausdruck von Evaluationen der Funktionsweise des Geräts funktionalisiert werden, verweist auf die besondere Bedeutung dieser Kategorie für die Untersuchung sprachlicher Praktiken. Dies gilt insbesondere in Verbindung mit den Befunden und offenen Fragen, die aus den Beispielen (7) und (8) zu Anthropomorphisierungen geschlussfolgert werden konnten. Weitere Untersuchungen und umfangreicheres Datenmaterial müssen zeigen, welche

14 Diese wird nach Zifonun et al. (1997, S. 405) „verwendet, um in einer Situation möglicher oder faktischer Divergenz, eines Reparaturbedarfs oder eines riskanten Übergangs eine kontinuierliche Fortsetzung des Diskurses zu ermöglichen“.

anderen sprachlichen Strategien (musterhaft) gebraucht werden, um erfüllte und enttäuschte Erwartungen an Sprachassistenten zum Ausdruck zu bringen und interaktional Evaluationen auszuhandeln. Inwiefern kann z.B. die mündliche Bedienbarkeit der Systeme, die auf der gleichen medialen Ebene liegt, wie die Interaktion mit ko-präsenten SprecherInnen, zu einer doppelten interaktional-technisierten Funktionalisierung von Äußerungen führen? Manifestiert sich diese in Beispiel (10) präsentierte Praktik als Routine?

Darüber hinaus zeigt das Beispiel auch, dass die Funktionsweise des Sprachassistenten gemeinsam erprobt wird. So gibt Josefine dem zu diesem Zeitpunkt „aktiven Bediener" des Systems mehrfach Hinweise zur Bedienweise (Z. 021, Z. 034). Diese interaktionale Dynamik konnte bereits in einer Studie von Pitsch et al. (2017, S. 395) untersucht werden. Entscheidend ist dabei in der Mensch-Maschine-Interaktion, dass das Assistenzsystem diese Phase der gesteigerten Dialogizität zwischen Mensch und Mensch als Stille auffasst und somit ein zweiter dialogischer Raum interaktional hergestellt wird, in dem der Sprachassistent nicht angesprochen wird. Zudem werden hier interaktional Rollen eines „Bedieners" und eines „Ko-Bedieners" ausgehandelt, die sich über die gesamte Sequenz verfestigen. Bemerkenswert ist dabei, dass sich der Dialog sowohl in der häuslichen Umgebung als auch mit dem Assistenzsystem von Josefine ereignet; dennoch übernimmt Josefine lediglich die Rolle einer Ko-Bedienerin, während Markus die Sprachbefehle artikuliert. Welche interaktionalen Dynamiken und welche sprachlichen Ausdrucksformen tragen zu einer Aushandlung dieser Rollen bei und wie verfestigen sie sich und werden durch die Beteiligten immer wieder neu hergestellt? Wie fixiert oder fluide sind diese Rollen über einen längeren Zeitraum hinweg? Diese Fragen verweisen auch auf die Bedeutung eines Sprachassistenzsystems innerhalb der sozialräumlichen Struktur eines privaten Haushalts mit zahlreichen Anknüpfungspunkten für interdisziplinäre Fragestellungen.

Fazit

Viele der in der Pilotstudie angewendeten Verfahren haben sich bewährt. So zeigt sich, dass der Zugang über private Haushalte und das „Schneeballsystem" schnelle und zufriedenstellende Ergebnisse liefern und die Situationstypen Aufschluss über die im Projekt aufgeworfenen Fragen geben können. Insofern kann konstatiert werden, dass die Datenerhebung so weitergeführt werden kann. Es lassen sich dennoch aus den Erkenntnissen der Pilotstudie einige Anpassungen ableiten, um die Ergebnisse der Datenerhebung zu verbessern: So sollte der Suchradius außerhalb des privaten Umfelds erweitert werden. Eine Möglichkeit, auf die auch über die Pilotstudie hinaus zurückgegriffen werden kann, ist die Ansprache von Studierenden und die Datenerhebung im Rahmen des oben bereits angesprochenen Seminars. Durch die Arbeit mit dem sog. Conditional Voice Recorder, durch bessere, auch schriftliche, Instruktionen und durch die Anwesenheit der ForscherInnen im Vorfeld der Video-Erhebungssituationen sollen die Qualität der Daten verbessert werden. Ferner lassen sich Datenschutzbedenken der TeilnehmerInnen reduzieren.

Die erhobenen und noch zu erhebenden Daten sind (nicht nur) linguistisch äußerst interessant: Es zeigen sich sprachliche Praktiken, die durch die Integration artifizieller Mündlichkeit in die Interaktion verändert werden oder neu entstehen. Dabei sind nicht nur emergente Praktiken zu beobachten, sondern Voice-User-Interfaces transformieren in der Interaktion offenbar auch bestehende sprachliche Praktiken. Diese Thesen konnten im Rahmen der Pilotstudie bestätigt werden. Anknüpfungspunkte für die Untersuchung und die Analysekategorien bilden entsprechend die näheren Untersuchungen von Kooperationsproblemen und dazugehörige Problemlösungsstrategien, der Einsatz und die Funktionalisierung von Anthropomorphisierung, die interaktiv ausgehandelte Nutzung sowie die komplexe inkrementelle Verflechtung von Mensch-Maschine-Interaktionen mit Mensch-Mensch-Interaktionen.

Forschungspraktisch ist entsprechend nun die Identifikation geeigneter Haushalte durch die beschriebenen Verfahren sowie die Vorbereitung passender Instruktionen geboten, um die vielversprechenden Daten gewinnen und sowohl linguistisch wie auch im anderen Teilbereich von B06 mediensoziologisch auszuleuchten.

Quellen

Bergmann, Jörg (2001): Das Konzept der Konversationsanalyse. In: Klaus Brinker, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 2. Halbband. Berlin/New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft/HSK, 16.2), S. 919–927.

Besthorn, Marit; Gerwinski, Jan; Habscheid, Stephan (2018): Methodik I: Erhebung, Aufbereitung, Archivierung, Datenschutz, gesprächslinguistische Auswertung und praxeologische Theoriebildung. In: Jan Gerwinski, Stephan Habscheid und Erika Linz (Hg.): Theater im Gespräch. Sprachliche Publikumspraktiken in der Theaterpause. Berlin/Boston: de Gruyter, S. 71–104.

Böckmann, Barbara; Meer, Dorothee; Mohn, Michelle; Och, Anastasia-Patricia; Paltrinieri, Ilaria; Renelt, Alina et al. (2019): Multimodale Produktbewertungen in Videos von Influencerinnen auf YouTube: Zur

parainteraktiven Konstruktion von Warenwelten. In: Zeitschrift für Angewandte Linguistik 70, S. 139–171.
Dasgupta, Ritwik (2018): Voice User Interface Design. Moving from GUI to Mixed Modal Interaction. New York: Apress.
Deppermann, Arnulf (2000): Ethnographische Gesprächsanalyse: Zu Nutzen und Notwendigkeit von Ethnographie für die Konversationsanalyse. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 1, S. 96–124. Online verfügbar unter www.gespraechsforschung-ozs.de.
Deppermann, Arnulf (2008): Gespräche analysieren. Eine Einführung. 4. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
Deppermann, Arnulf (2013): Zur Einführung: Was ist eine „interaktionale Linguistik des Verstehens"? In: Deutsche Sprache 13, S. 1–5.
Ehlich, Konrad (2007): Sprache und sprachliches Handeln. Band 1: Pragmatik und Sprechakttheorie. Berlin: de Gruyter.
Epley, Nicholas; Waytz, Adam; Cacioppo, John T. (2007): On Seeing Human: A Three-Factor Theory of Anthropomorphism. In: Psychological Review 114 (4), S. 864–886.
Fiehler, Reinhard (2002): How to Do Emotions With Words: Emotionality in Conversation. In: Susan R. Fussell (Hg.): The Verbal Communication of Emotions. Interdisciplinary Perspectives. London: Lawrence Erlbaum, S. 79–106.
Garfinkel, Harold (1967): Studies in Ethnomethodology. Cambridge: Polity.
Gerwinski, Jan; Linz, Erika (2018): Methodik II: Beobachterparadoxon – die Aufnahmesituation im Gespräch. Unter Mitarbeit von Marit Besthorn. In: Jan Gerwinski, Stephan Habscheid und Erika Linz (Hg.): Theater im Gespräch. Sprachliche Publikumspraktiken in der Theaterpause. Berlin/Boston: de Gruyter, S. 105–163.
Golato, A. (2010): Marking understanding versus receipting information in talk: Achso. and ach in German interaction. In: Discourse Studies 12 (2), S. 147–176.
Günthner, Susanne (2016): Praktiken erhöhter Dialogizität: onymische Anredeformen als Gesten personifizierter Zuwendung. In: ZGL 44 (3), S. 406–436.
Habscheid, Stephan (2016): Handeln in Praxis. Hinter- und Untergründe situierter sprachlicher Bedeutungskonstitution. In: Arnulf Deppermann, Helmuth Feilke und Angelika Linke (Hg.): Sprachliche und kommunikative Praktiken. Berlin u.a.: de Gruyter, S. 127–151.
Habscheid, Stephan; Hrncal, Christine; Lüssem, Jens; Wieching, Rainer; Carros, Felix; Wulf, Volker (2019): Robotics and Emotion. In: Nicole Shea und Emmanuel Kattan (Hg.): Europe now. Special Feature „Anxiety Culture" of „Europe now". Columbia: Council for European Studies at Columbia University. Online verfügbar unter https://www.europenowjournal.org/2018/07/01/robotics-and-emotion/.
Have, Paul ten (1999): Doing Conversation Analysis. A Practical Guide. London: Sage.
Hennig, Martin; Hauptmann, Kilian (2019): Alexa, optimier mich! KI-Fiktionen digitaler Assistenzsysteme in der Werbung. In: Zeitschrift für Medienwissenschaft 11 (21), S. 86–94.
Imo, Wolfgang (2009): Konstruktion oder Funktion? Erkenntnisprozessmarker (change-of-state token) im Deutschen. In: Jörg Bücker und Susanne Günther (Hg.): Grammatik im Gespräch. Konstruktionen der Selbst- und Fremdpositionierung. Berlin u.a.: de Gruyter, S. 57–86.
Kotthoff, Helga (2003): Was heißt eigentlich doing gender? Differenzierungen im Feld von Interaktion und Geschlecht. In: FZG–Freiburger Zeitschrift für Geschlechter Studien 9 (1), 125-161.
Krüger, Julia; Wahl, Mathias; Frommer, Jörg (2018): „es is komisch es is keen mensch" – Zuschreibungen gegenüber individualisierten technischen Assistenzsystemen. Eine Interviewstudie zum Nutzer/innenerleben in der Mensch-Computer-Interaktion. In: ZQF 19 (1-2), S. 233–250.
Krummheuer, Antonia (2010): Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts. Stuttgart: Lucius&Lucius.
Kupetz, Maxi (2014): Empathy displays as interactional achievements – Multimodal and sequential aspects. In: Journal of Pragmatics 61, S. 4–34. DOI: 10.1016/j.pragma.2013.11.006.
Kupetz, Maxi (2015): Empathie im Gespräch. Tübingen: Stauffenburg (Stauffenburg Linguistik, Band 88).
Lau, Josephine; Zimmerman, Benjamin; Schaub, Florian (2018): Alexa, Are You Listening? In: Proceedings of the ACM on Human-Computer Interaction 2, S. 1–31. DOI: 10.1145/3274371.
Linke, Angelika; Nussbaumer, Markus; Portmann-Tselikas, Paul R.; Willi, Urs (1996): Studienbuch Linguistik. 3., unveränd. Aufl. Tübingen: Niemeyer.
Lotze, Netaya (2016): Chatbots. Eine linguistische Analyse. Frankfurt a.M.: Peter Lang.
Marquardt, Manuela (2017): Anthropomorphisierung in der Mensch-Roboter-Interaktionsforschung: theoretische Zugänge und soziologisches Anschlusspotenzial. In: Working Papers kultur- und techniksoziologische Studien 10 (1), S. 4–43. Online verfügbar unter https://nbn-resolving.org/urn:nbn:de:0168-ssoar-57037-3.
Mondada, Lorenza (2014): Conventions for multimodal transcription. In: Ecole thématique CNRS. MAINLY - MultimodAL (INter)actions LYon: the construction and organisation of social actions. Online verfügbar unter https://mainly.sciencesconf.org/conference/mainly/pages/Mondada2013_conv_multimodality_copie.pdf.
Pearl, Cathy (2016): Designing Voice User Interfaces. Principles of Conversational Experiences. Sebastopol: O'Reilly Media. Online verfügbar unter http://proquest.tech.safaribooksonline.de/9781491955406.
Pitsch, Karola (2015): Ko-Konstruktion in der Mensch-Roboter-Interaktion. Kontingenz, Erwartungen und Routinen in der Eröffnung. In: Ulrich Dausendschön-Gay, Elisabeth Gülich und Ulrich Krafft (Hg.): Ko-Konstruktionen in der Interaktion. Die gemeinsame Arbeit an Äußerungen und anderen sozialen Ereignissen. Bielefeld: Transcript, S. 229–257.
Pitsch, Karola; Gehle, Raphaela; Dankert, Timo; Wrede, Sebastian (2017): Interactional Dynamics in User Groups. In: Britta Wrede (Hg.): Proceedings of the 5th International Conference on Human Agent Inter-

action. Bielefeld, Germany, 10/17/2017 - 10/20/2017. New York: ACM Press, S. 393–397.
Porcheron, Martin; Fischer, Joel E.; Reeves, Stuart; Sharples, Sarah (2018): Voice Interfaces in Everyday Life. In: Regan Mandryk, Mark Hancock, Mark Perry und Anna Cox (Hg.): Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Montreal QC, Canada, 21.04.2018 - 26.04.2018. New York: ACM Press, S. 1–12.
Potter, Jonathan; Wetherell, Margaret (1987): Discourse and Social Psychology. Beyond Attitudes and Behaviour. London: Sage.
Sacks, Harvey (1995): Lectures on Conversation. Volume I. Oxford: Blackwell.
Schatzki, Theodore R. (2002): The Site of the Social. A Philosophical Account of the Constitution of Social Life and Change. University Park: Pennsylvania State University Press.
Schegloff, Emanuel (1997): Practices and Actions: Boundary Cases of Other-Initiated Repair. In: Discourse Processes 23, S. 499–545.
Schegloff, Emanuel (2012): Interaktion: Infrastruktur für soziale Institutionen, natürliche ökologische Nische der Sprache und Arena, in der Kultur aufgeführt wird. In: Ruth Ayaß und Christian Meyer (Hg.): Sozialität in Slow Motion. Theoretische und empirische Perspektiven. Wiesbaden: Springer VS, S. 245–268.
Schenkein, Jim (Hg.) (1978): Studies in the Organization of Conversational Interaction. New York: Academic Press.
Schlinkmann, Eva; Hesse, Mareike (2018): Settings und Sampling. In: Jan Gerwinski, Stephan Habscheid und Erika Linz (Hg.): Theater im Gespräch. Sprachliche Publikumspraktiken in der Theaterpause. Berlin/Boston: de Gruyter, S. 17–70.
Schüttpelz, Erhard; Meyer, Christian (2017): Ein Glossar zur Praxistheorie. „Siegener Version" (Frühjahr 2017). In: Navigationen 17 (1), S. 155–163.
Selting, Margret (2016): Praktiken des Sprechens und Interagierens im Gespräch aus Sicht von Konversationsanalyse und Interaktionaler Linguistik. In: Arnulf Deppermann, Helmuth Feilke und Angelika Linke (Hg.): Sprachliche und kommunikative Praktiken. Berlin u.a.: de Gruyter, S. 27–56.
Selting, Margret; Auer, Peter; Barth-Weingarten, Dagmar; Bergmann, Jörg; Bergmann, Pia; Birkner, Karin et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung. Online-Zeitschrift zur verbalen Interaktion 10, S. 353–402. Online verfügbar unter www.gespraechsforschung-osz.de.
Selting, Margret; Couper-Kuhlen, Elizabeth (2001): Forschungsprogramm ‚Interaktionale Linguistik'. In: Linguistische Berichte 187, S. 257–287.
West, Emily (2019): Amazon: Surveillance as a Service. In: Surveillance & Society 17(1/2), S. 27–33. Online verfügbar unter https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/13008/8472.
West, Candace; Zimmermann, Don (1989): Doing Gender. In: Gender & Society 2, S. 121–151.
Wissenschaftliche Dienste des Deutschen Bundestages (2019): Zulässigkeit der Transkribierung und Auswertung von Mitschnitten der Sprachsoftware „Alexa" durch Amazon, 29.05.2019. Online verfügbar unter https://www.bundestag.de/resource/blob/650728/3f72e6abc1c524961e5809002fe20f21/WD-10-032-19-pdf-data.pdf.
Zifonun, Gisela; Hoffmann, Ludger; Strecker, Bruno (1997): Grammatik der deutschen Sprache. Unter Mitarbeit von Joachim Ballweg, Ursula Brauße, Eva Breindl, Ulrich Engel, Helmut Frosch, Ursula Hoberg und Klaus Vorderwülbecke. 3 Bände. Berlin u.a.: de Gruyter.

Anhang

GAT 2-Transkriptionskonventionen (Selting et al. 2009)

Sequenzielle Struktur/Verlaufsstruktur

[ ] [ ]	Überlappungen und Simultansprechen

Ein- und Ausatmen

°h/h°	Ein- bzw. Ausatmen von ca. 0.2-0.5 Sek. Dauer

Pausen

(.)	Mikropause, geschätzt, bis ca. 0.2 Sek. Dauer
(-)	kurze geschätzte Pause von ca. 0.2-0.5 Sek. Dauer
(--)	mittlere geschätzte Pause von ca. 0.5-0.8 Sek. Dauer
(0.5)	gemessene Pausen von ca. 0.5 bzw. 2.0 Sek. Dauer

Sonstige segmentale Konventionen

und_äh	Verschleifungen innerhalb von Einheiten
äh öh äm	Verzögerungssignale, sog. „gefüllte Pausen"
:	Dehnung, Längung, um ca. 0.2-0.5 Sek.
::	Dehnung, Längung, um ca. 0.5-0.8 Sek.
:::	Dehnung, Längung, um ca. 0.8-1.0 Sek.

Lachen und Weinen

haha hehe hihi	silbisches Lachen
((lacht))((weint))	Beschreibung des Lachens
<<lachend> >	Lachpartikeln in der Rede, mit Reichweite
<<:-)> soo>	„smile voice"

Rezeptionssignale

hm ja nein nee	einsilbige Signale
hm_hm ja_a nei_ein nee_e	zweisilbige Signale
?hm?hm	mit Glottalverschlüssen, meistens verneinend

Sonstige Konventionen

((hustet))	para- und außersprachliche Handlungen u. Ereignisse
<<hustend> >	sprachbegleitende para- und außersprachliche Handlungen und Ereignisse mit Reichweite
( )	unverständliche Passage ohne weitere Angaben
(solche)	vermuteter Wortlaut
(also/alo)	mögliche Alternativen
((unverständlich, ca. 3 Sek))	unverständliche Passage mit Angabe der Dauer
((...))	Auslassung im Transkript

Akzentuierung

akZENT	Fokusakzent
ak!ZENT!	extra starker Akzent

Tonhöhenbewegung am Ende von Intonationsphrasen

?	hoch steigend
,	mittel steigend
–	gleichbleibend
;	mittel fallend
.	tief fallend

Transkriptionskonventionen für multimodale Transkripte (Mondada 2014)

```
* *      Gestures and descriptions of embodied actions are delimited between
+ +      two identical symbols (one symbol per participant)
∆ ∆      and are synchronized with correspondent stretches of talk.
*--->    The action described continues across subsequent lines
---->*   until the same symbol is reached.
----     Action's apex is reached and maintained.
 ric     Participant doing the embodied action is identified when (s)he is not the speaker.
```